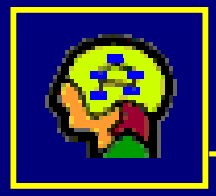
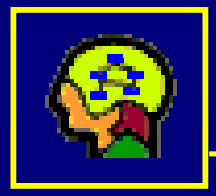


Integrated Intelligent Support for Knowledge Capture, Refinement and Sharing†

David B. Leake, Indiana University
Alberto Cañas, IHMC, University of West Florida

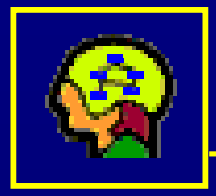
October 24, 2002

†Supported by the NASA Intelligent Systems Program under award No NCC 2-1216.



Overview

- Project Goal
- Concept maps (Cmaps)
- The CmapTools
- Supporting the concept mapping process
  - The case-based suggester
  - Mining the web for relevant concepts
  - Mining the web for related topics
- Conclusions



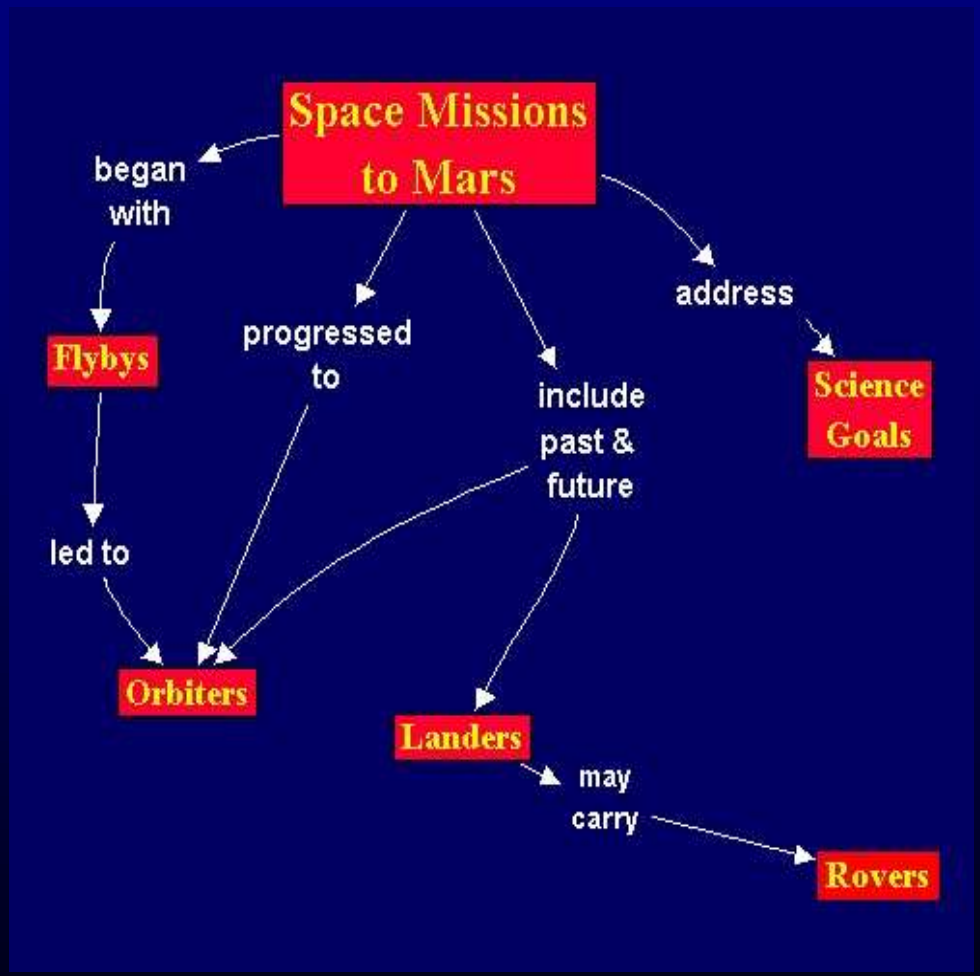
Project Overview

- Goal: To develop intelligent support tools for knowledge modeling, to empower experts to directly build, share, compare and refine knowledge models
- Primary method: Supporting concept mapping

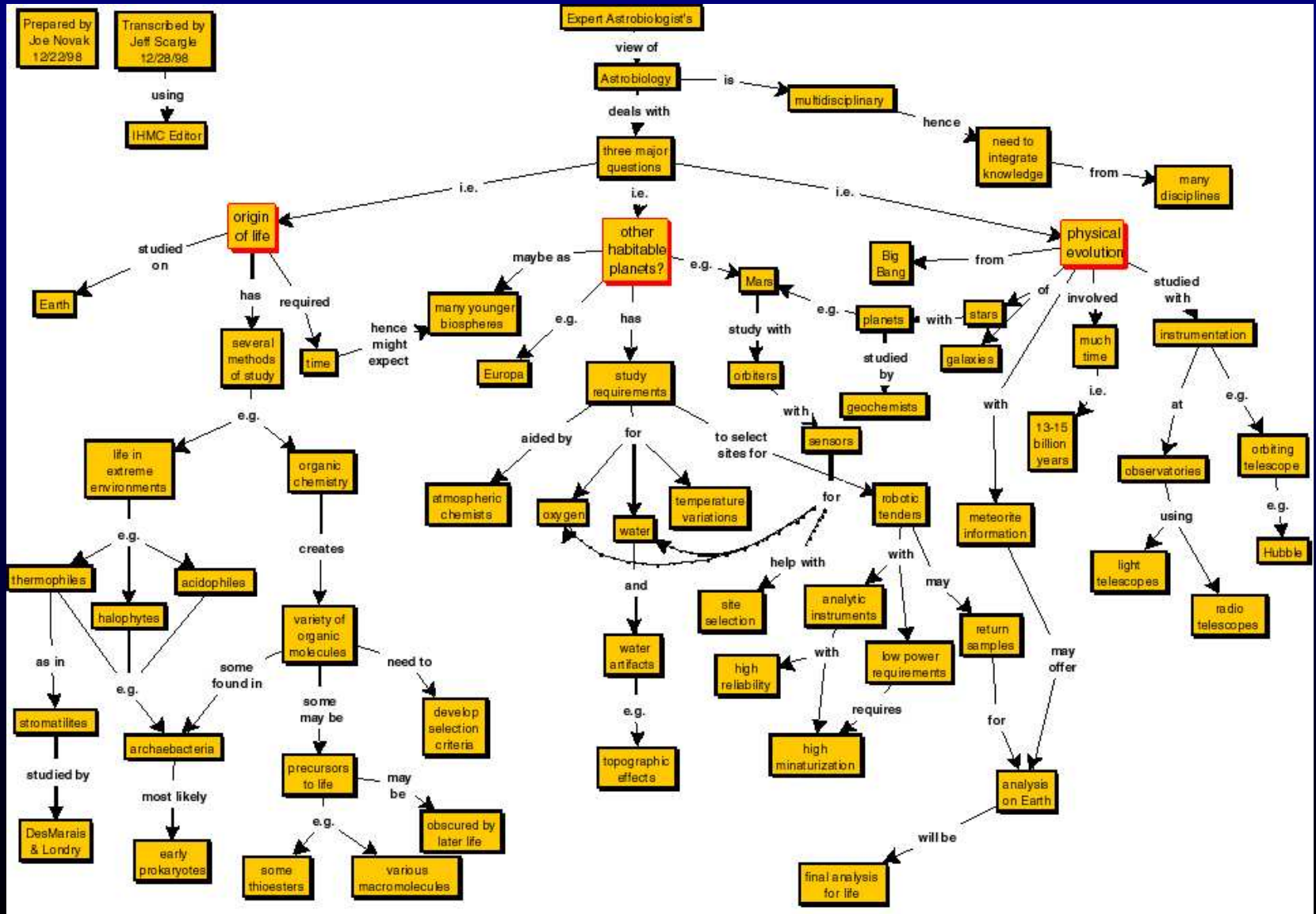
Concept Maps

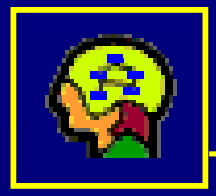
- Developed in an educational setting by Novak (1977)
- Used as the primary language for description and communication of concepts within Assimilation theory (Ausubel)
- A Concept map is a graphical display of concept names connected by directed arcs encoding propositions in the form of simplified sentences
- In educational settings, concept mapping techniques have aided people of every age to examine many fields of knowledge

A Concept Map Fragment



A Concept Map for the Astrobiology Institute



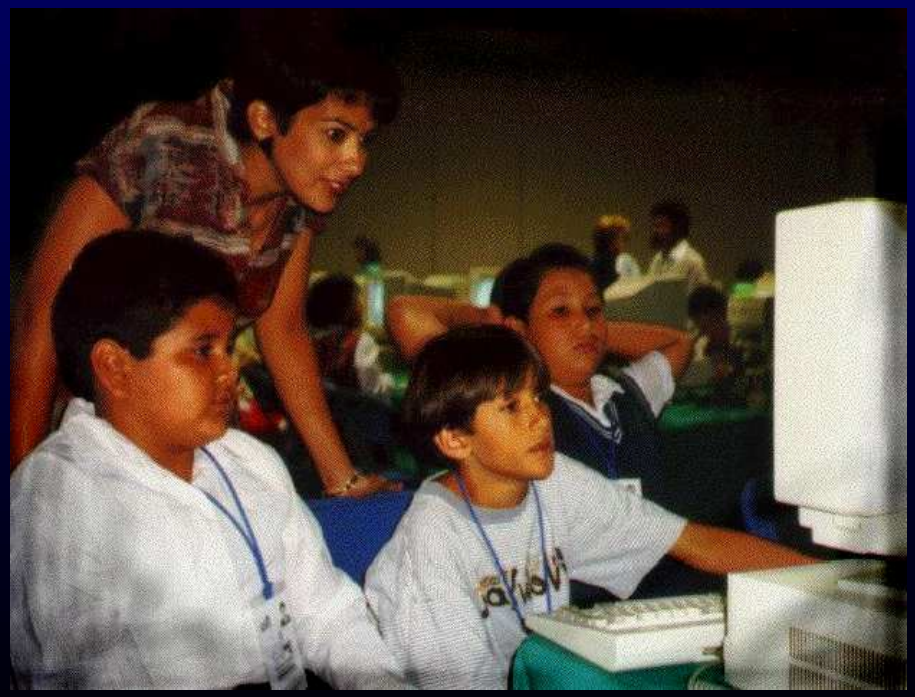


From an AI Perspective

- Concept maps seem similar to semantic networks and conceptual graphs
- …but concept maps:
  - are not “knowledge representations” in a computational or logical sense
  - are much more loosely defined, with no firm syntactic rules and no formal rules of interpretation or semantics.
- Concept maps are a …
  - pedagogic device for use by humans rather than
  - … a formal device for use by reasoning engines

Goals of the CmapTools Project at the Institute for Human and Machine Cognition

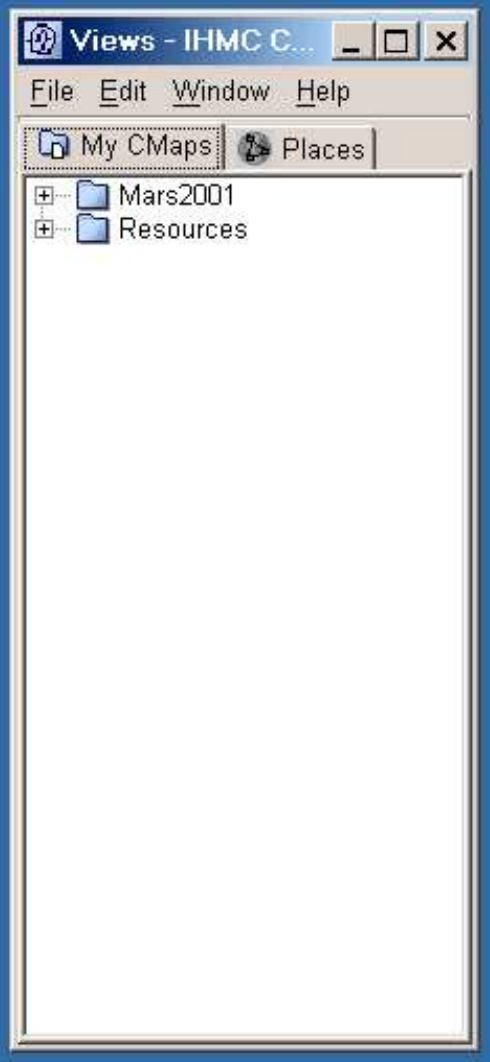
- Provide a methodological and technological foundation for distributed collaboration



- Allow students, scientists, and users in general to:
  - construct,
  - experiment with,
  - navigate,
  - criticize,
  - and share

  ... knowledge models about specific domains using concept maps

Generating a Concept Map using CmapTools

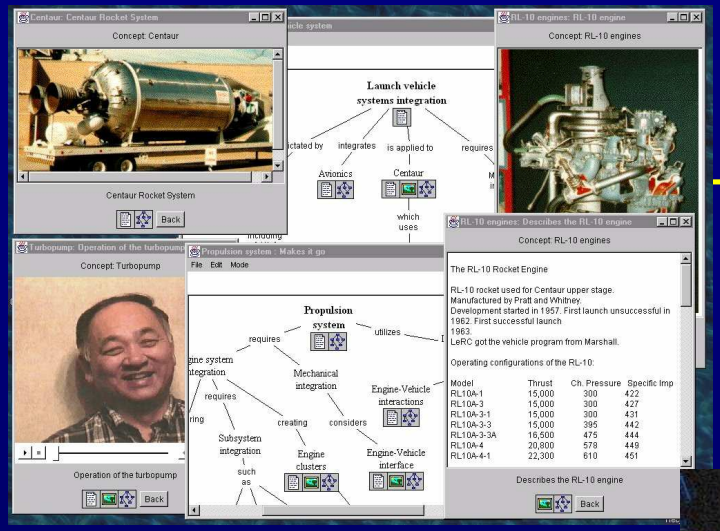


Sample Application: "Return to Mars 2001" Project

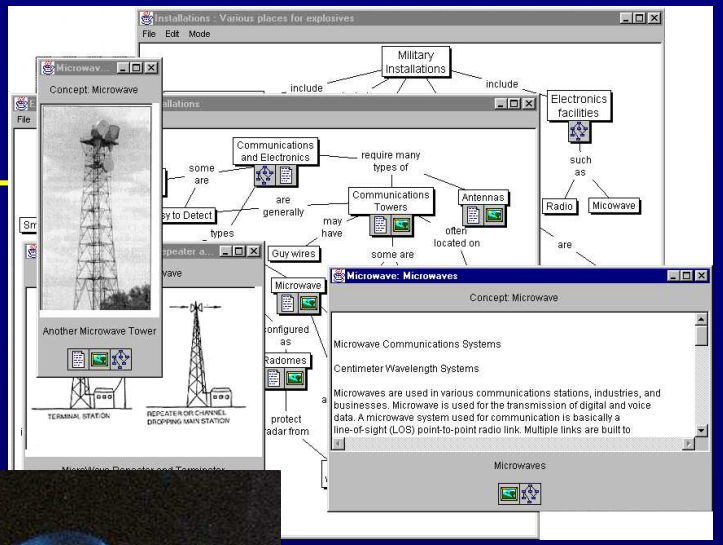
NASA Ames Research Center

- Capture NASA's Mars expert's domain knowledge into browsable concept maps that will be made available to high-school students, and to the public, on a CD and on the Web

- Concept Maps developed by Dr. Geoff Briggs, the Director of the Center for Mars Exploration at NASA Ames, in collaboration with other scientists

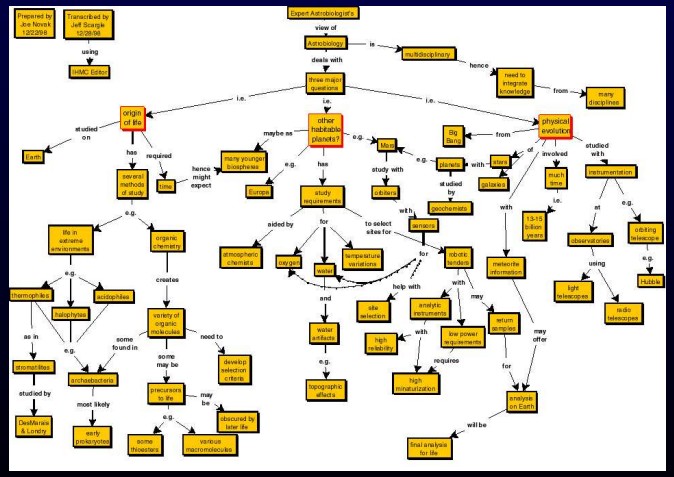


Explanation Component Expert Systems



Corporate Memory
NASA Lewis

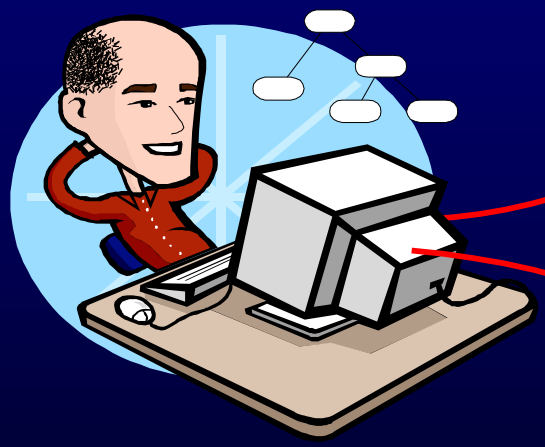
Perfomance Support System with Embedded Training;
Distance Learning / US Navy



Knowledge sharing among scientists Astrobiology Institute / NASA

Thailand Artisan's Knowledge

Aiding Users Constructing Concept Maps



User/Expert constructing a Knowledge Model

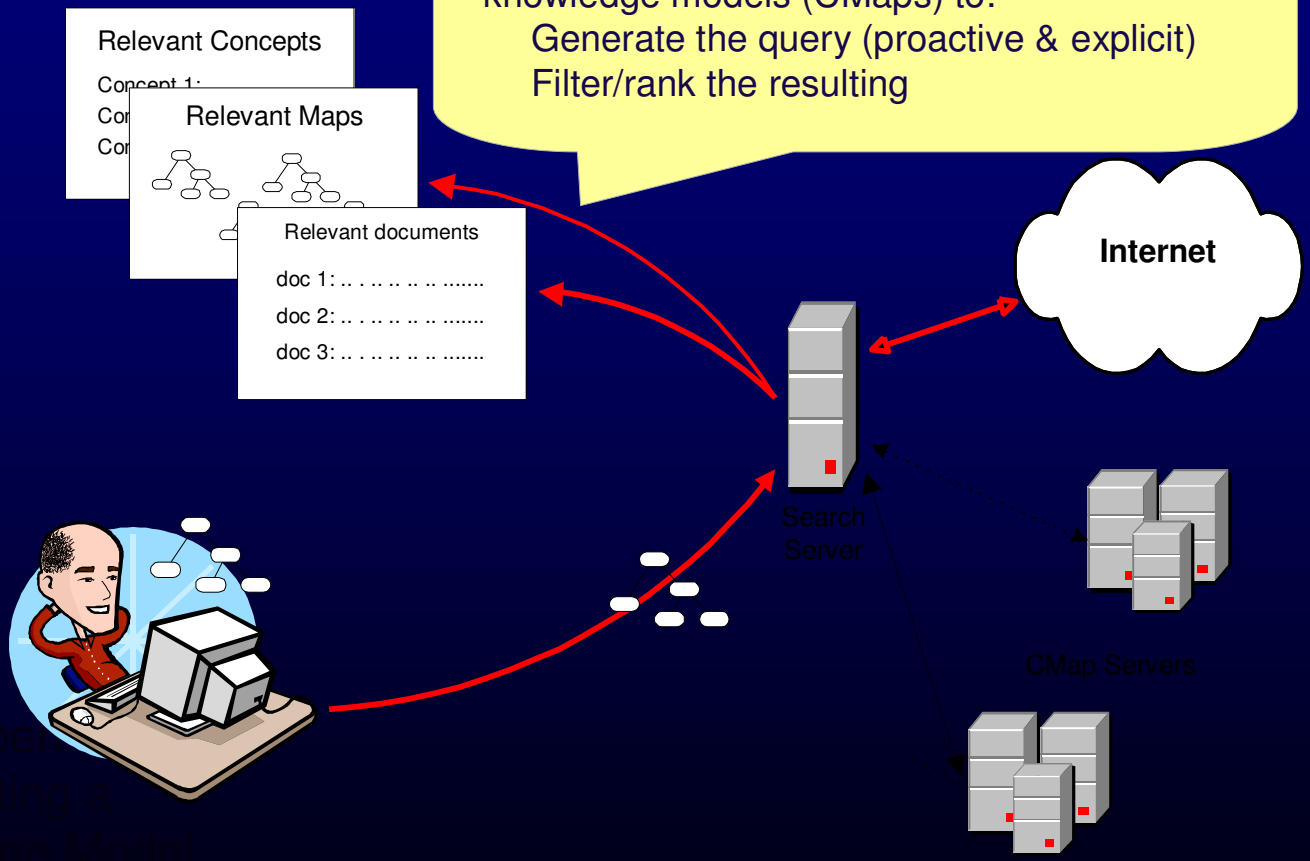
Goals for “Suggestions” window:

- Relevant Concept Maps
- Relevant Web pages
- Feedback regarding “goodness” of the Concept Map
- Suggested propositions
- Suggested concepts
- Suggested topics
- Questions

Aiding Users in the Construction of CMaps

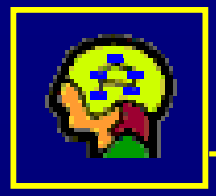
Take advantage of graphical representation of knowledge models (CMaps) to:

- Generate the query (proactive & explicit)
- Filter/rank the resulting



Current Research Focuses

1. Case-based suggester
   - Initial results for retrieval quality shown last review
   - Index methods developed for more efficient retrieval; test implementation complete
2. Web-based mining for new concepts
   - Mining for specific concept suggestions to consider in new maps
   - Mining procedures implemented
   - Promising initial experimental results
3. Web-based mining for new topics
   - Developing methods to suggest topics for new maps



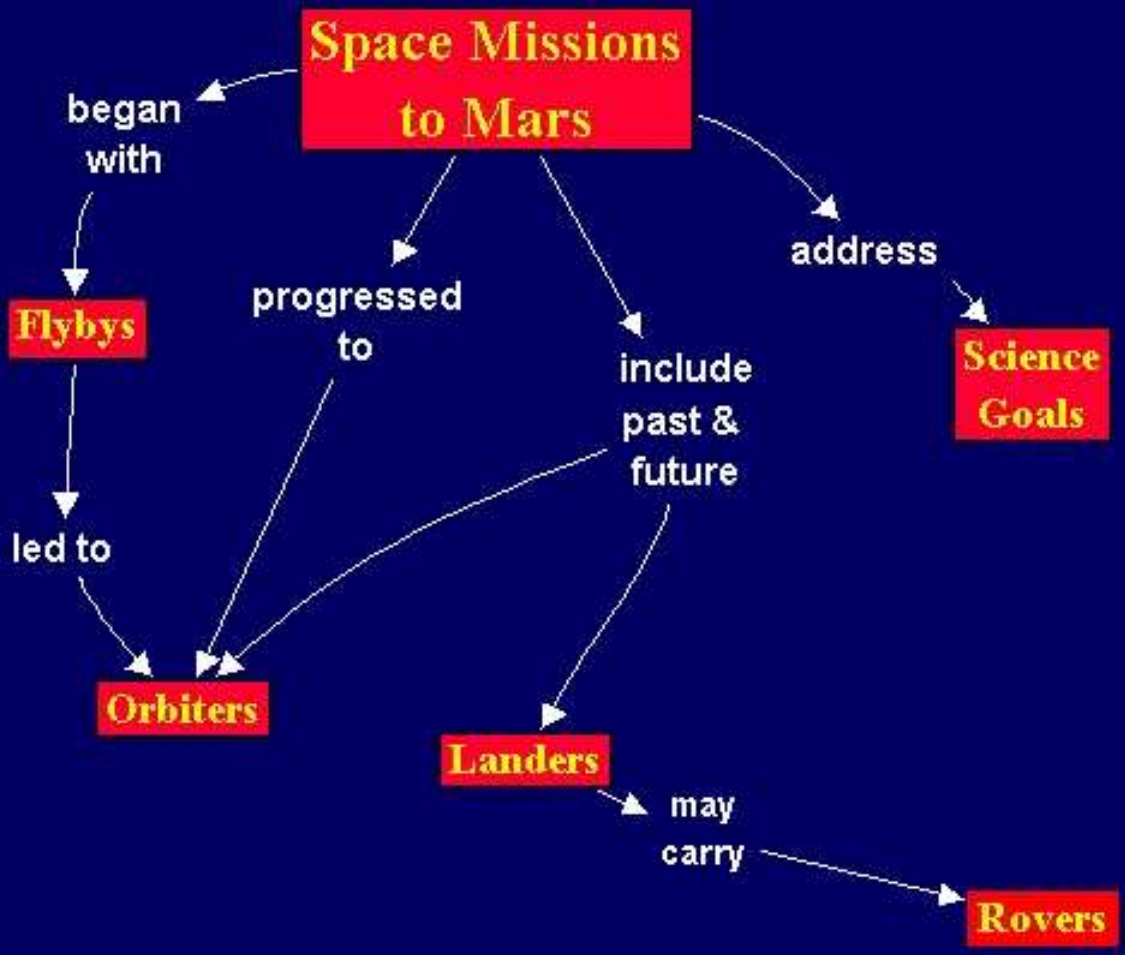
1. The CMap/CBR project

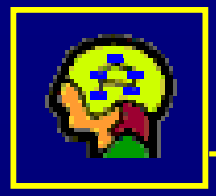
Case-based reasoning solves new problems by retrieving and adapting solutions to similar prior problems.

The case-based suggester retrieves similar prior concept maps to support:

- *Knowledge capture and refinement* by analogy to previous concept maps
- *Knowledge sharing:* Suggesting relevant concept maps from distributed local and remote libraries

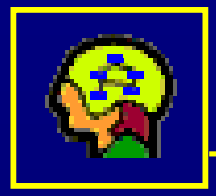
Generation of Concept Extension





Central Issues for applying CBR

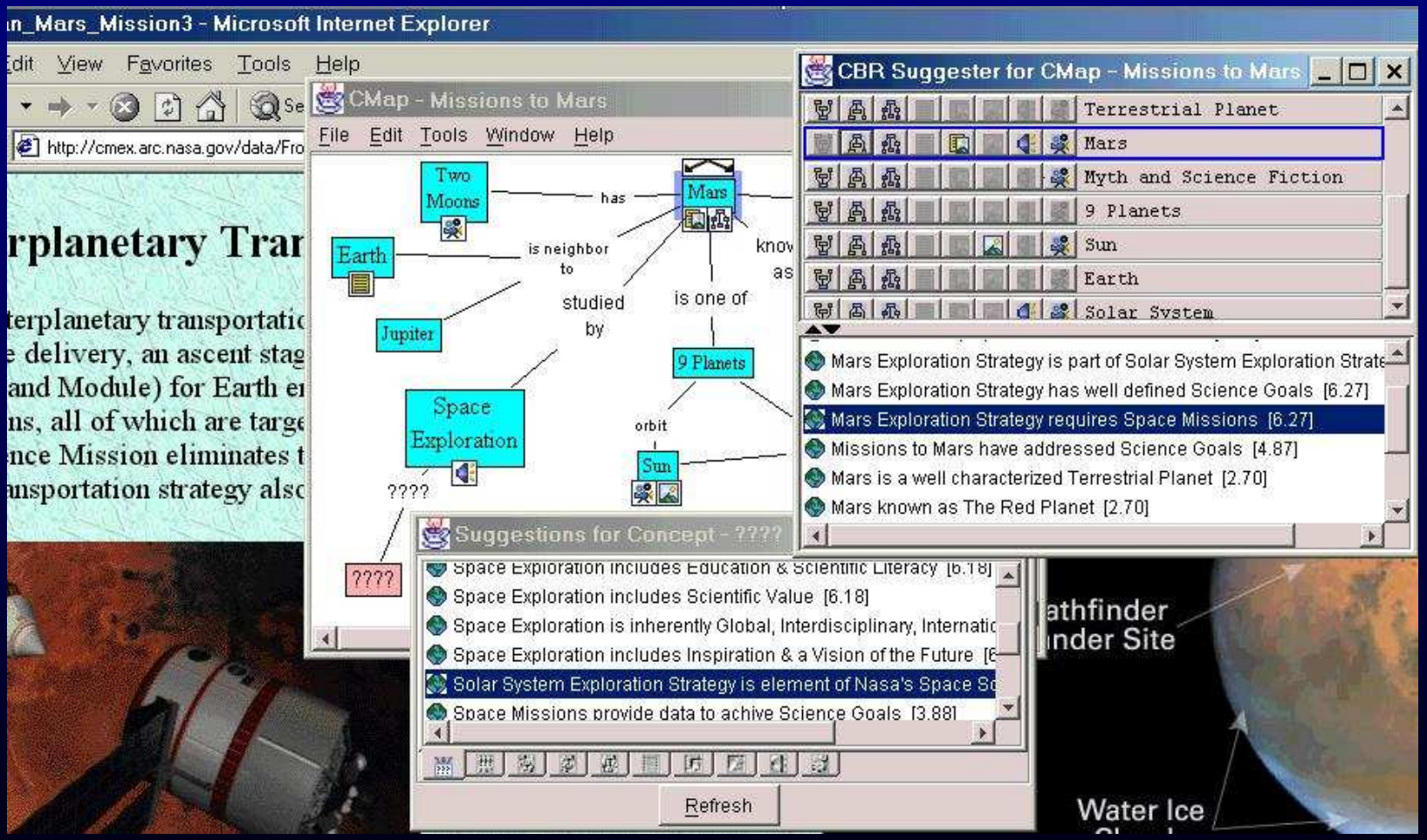
- The roles of content and structure in similarity assessment
- Assessing similarity and relevance for non-standardized representations
- Efficient use of structural information
- Exploiting contextual information



Progress to Date on CBR Suggester

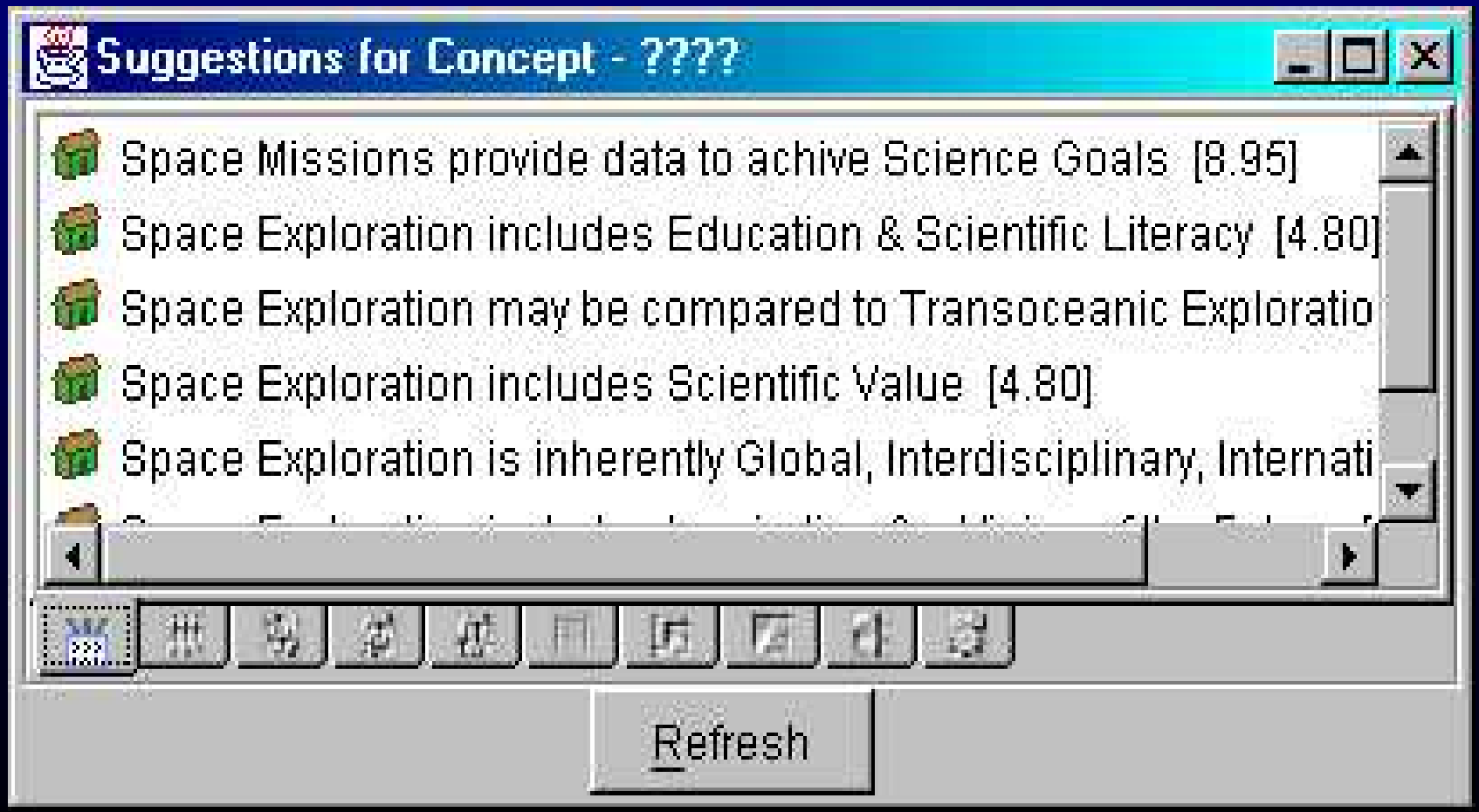
- Initial Suggester implemented and tested
- Primary focus is on suggesting concepts to link to a current concept
- Pilot study to compare Suggester's similarity rankings to human subjects'
  - Shows benefits compared to simple baseline
  - Suggests value of considering both global correlations and context of the concept to extend
- Current focuses:
  - Indexing to increase efficiency for scalability
  - Designing additional experiments

The CBR Suggester Window

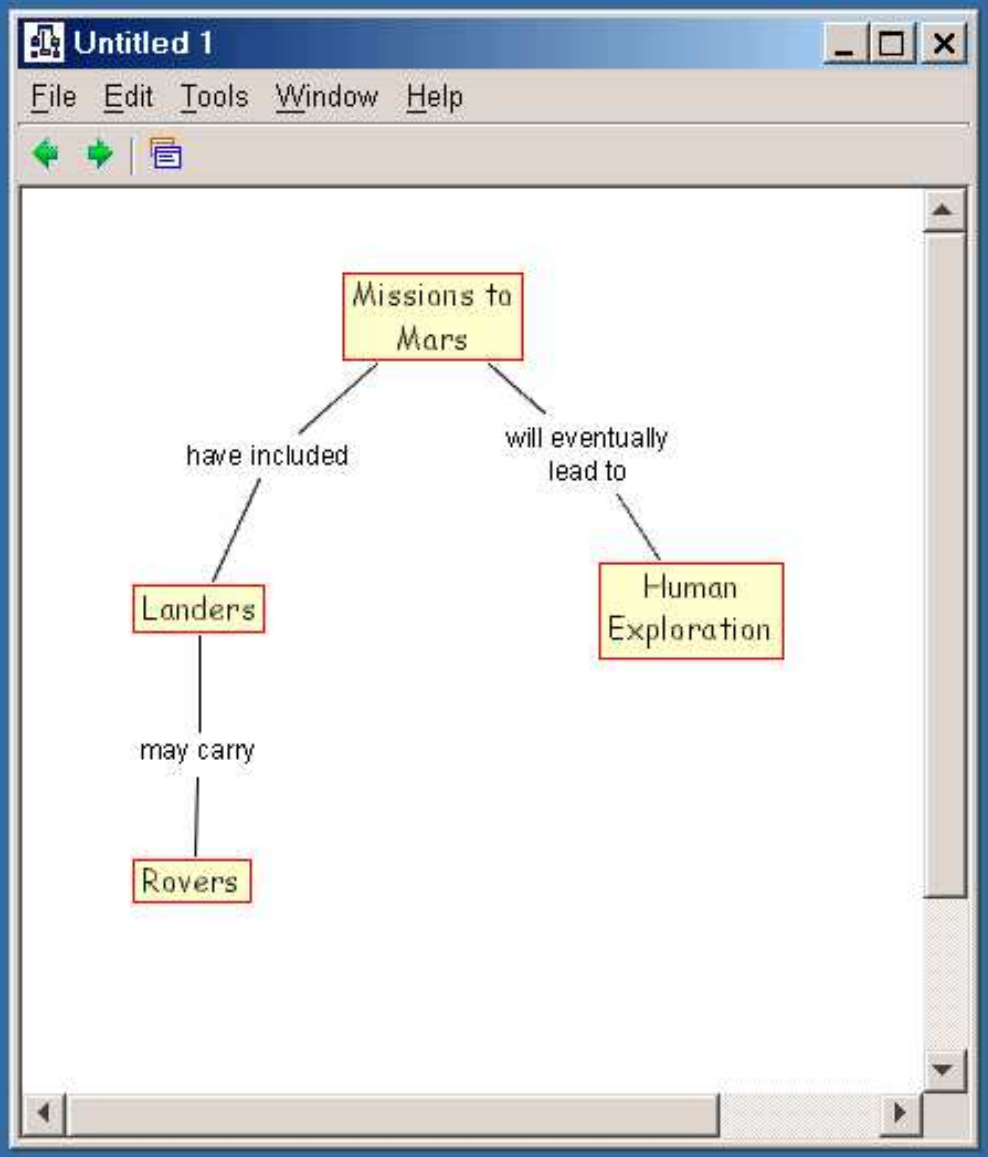


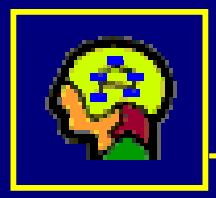
Suggestions for Extending "Space Missions to Mars"

Space Missions to Mars → ???? → ????



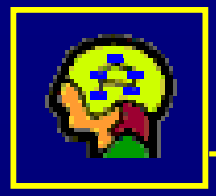
An Example in Use





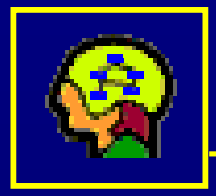
Refining the CBR Suggester: Clustering to Facilitate CMap Retrieval

1. Use Category Index to efficiently recognize a best-fit category for a concept map in progress.
2. Use more expensive comparisons to rank potential suggestions.



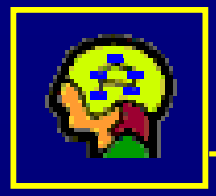
Category Indexing

- Grouping concept maps by topic helps focus search
- Deciding which maps to group is based on the cosine measure of a TFIDF representation tailored for concept maps (CMap-TFIDF).
- A category is a coherent set of maps.
- Agglomerative clustering of concept map representations generates a category hierarchy.



Clustering by CMap Document Model

- Vector model of keywords in CMap.
- Measures relevance of keywords to specify category. Uses:
  - TFIDF
  - number of outgoing and incoming links of a node
  - distance from root node
  - co-occurrence of neighboring keywords
- Document similarity measured by degree of overlap of shared keywords.

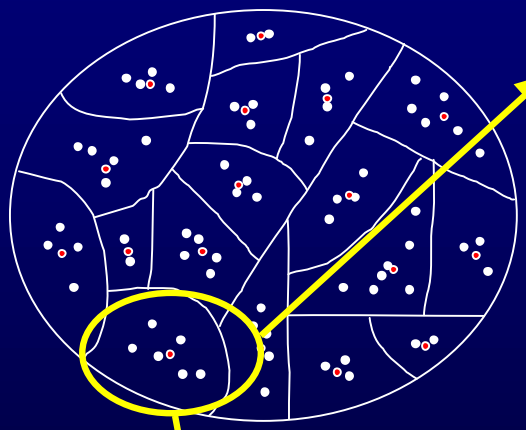


CMap Document Clustering

- Agglomerative greedy clustering algorithm.
- Merges repeatedly pairs of clusters whose centroid vectors are closest.
- Merging terminates if maximum number of clusters is reached or if cluster's similarity falls below threshold.
- Centroid vector of cluster represents average document similarity.

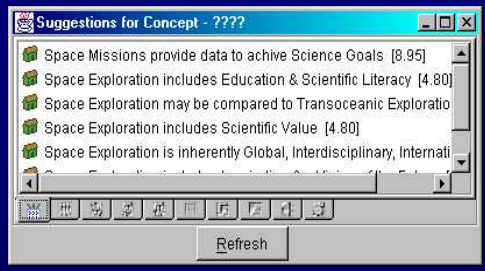
The Role of the Meta-Index

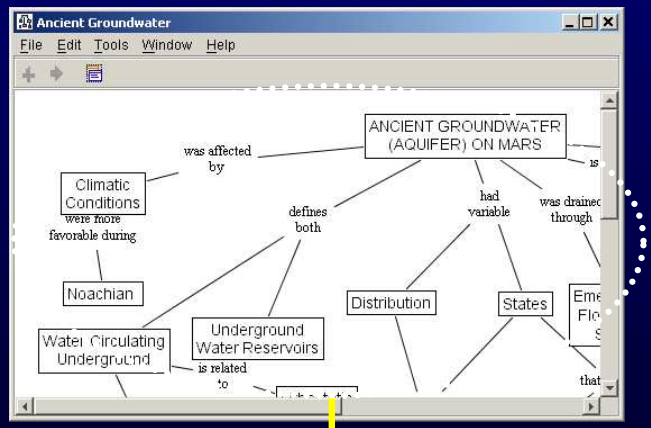
Meta-Index



3. Generate Suggestions for CMap using Case Library.

4. Present suggestions





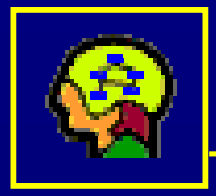
Case Library

2. Match closest categories.

1. Convert CMap to Vector Repre-sentation.

CBR Service on Index Server

CBR Client Suggestor

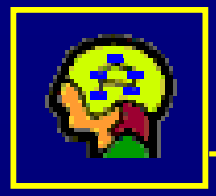


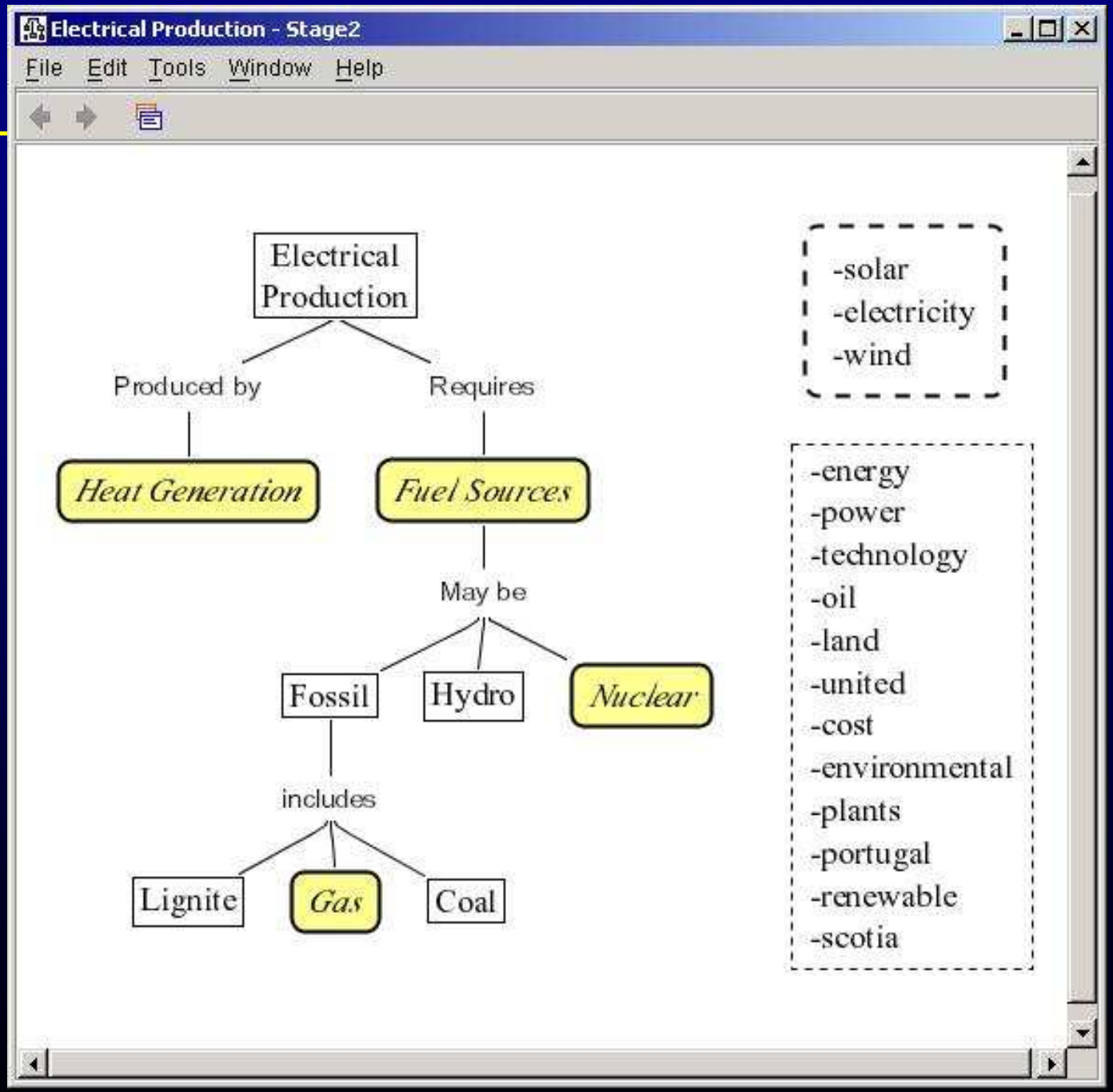
2. Mining the Web for Concept Suggestions

- Terms are extracted from concept labels
- Stopwords are deleted and short labels are retained
- Terms are input to metasearch engine
- Returned documents are ranked based on frequency of concepts in the map, and terms in close proximity are candidate suggestions.
- Suggestions are returned ranked by frequency

Experimental Evaluation
(Cañas, Carvalho & Arguedas in press)

- 7 subjects generated concept maps relevant to “how do we produce electricity”, without receiving suggestions
- Their sequence of concepts was recorded
- After the session, the web-based suggester was run on each incremental map for each subject.
- Results counted the suggestions that matched concepts actually used later in the construction process.
- Depending on the size of the suggestion list, the suggester presented on average 47%-69% of the concepts actually added to the map.




Electrical Production - Stage2
File Edit Tools Window Help
Electrical Production
Produced by
Requires
Heat Generation
Fuel Sources
May be
Fossil
Hydro
Nuclear
includes
Lignite
Gas
Coal
-solar
-electricity
-wind
-energy
-power
-technology
-oil
-land
-united
-cost
-environmental
-plants
-portugal
-renewable
-scotia

Experimental Results

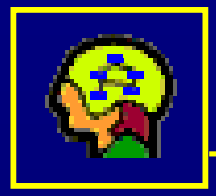
Final Concept Maps									
Subject	Total # of Unique Concepts	Number of Concepts in Map that were in a Previous Suggestion List							
		From List of 15 Suggestions		From List of 25 Suggestions		From List of 50 Suggestions		From the Complete List	
		#	%	#	%	#	%	#	%
1	15	9	60%	11	73%	12	80%	13	87%
2	16	6	38%	8	50%	10	63%	10	63%
3	16	9	56%	9	56%	12	75%	13	81%
4	13	3	23%	3	23%	3	23%	4	31%
5	28	7	25%	10	36%	13	46%	14	50%
6	17	12	71%	14	82%	15	88%	15	88%
7	13	7	54%	9	69%	10	77%	11	85%
Average	**16.86**	**7.57**	**47%**	**9.14**	**56%**	**10.71**	**65%**	**11.43**	**69%**

3. Web Mining for New Topics

- Goal to suggest novel but coherent topics to extends an in-progress concept map
- EXTENDER
  - Starts from a set of one or more concept maps and produces the first generation of artificial topics
  - Produces successive generations of artificial topics by:
    - Retrieving related web pages
    - Applying clustering to produce the next generation of artificial topics

Extender's Methods

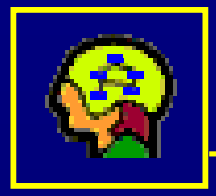
- *Topological Analysis* to summarize the roles of keywords in originating concept maps
- *Diversity/Focus factors* to guide exploration
- *Concept Cohesion approach* for combining terms into queries
- *Association clusters and weighting techniques* for recognizing relevant novel keywords



Artificial Topics Generated

(starting from a concept map on Planet Mars)

mars	climate	robotic	history
science	global	missions	climate
goals	change	human	geologic
exploring	changes	space	earth
nasa	environmental	web	change
space	water	exploration	geology
exploration	research	future	processes
missions	activities	system	global
mission	usgs	services	water
system	future	explore	changes
rover	national	nasa	life
future	amp	home	geological
earth	program	page	natural



Pilot Experiment

Setting:

- The Mars 2001 Project contains 100 concept maps created by experts from NASA
- Only one concept map from the Mars 2001 Project was used by EXTENDER to produce 27 artificial topics
- The cosine function was used to compute 2700 similarity measures between the 27 artificial topics and the 100 existing concept maps

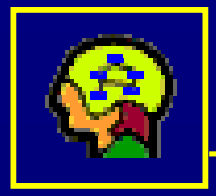
Results:

- The measures returned from comparing the originating concept map and each artificial topic were always below 0.40
- We obtained 47 similarity measures with a value over 0.50, where 7 were over 0.70

We can conclude that EXTENDER predicted novel artificial topics with content similar to that of concept maps created by experts

Conclusion

- Concept mapping provides a flexible method for empowering experts to construct rich knowledge models
- The CmapTools support distributed access and sharing
- The suggesters can provide
  - Automatic access to prior maps for suggestions and comparisons
  - Suggestions of concepts and topics from web documents, enabling drawing on a wider range of resources
- Next steps include: evaluation of indexing, extension and refinement of web-based suggester algorithms, suggester integration, and more extensive evaluations of alternative methods with human subjects.



Integrated Intelligent Support for Knowledge Capture, Refinement and Sharing†

David B. Leake, Indiana University
Alberto Cañas, IHMC, University of West Florida

October 24, 2002

†Supported by the NASA Intelligent Systems Program under award No NCC 2-1216.